LBL-25772

Submitted to *Building and Environment*

ON ESTIMATION OF MULTIZONE VENTILATION RATES FROM TRACER-GAS MEASUREMENTS

M. H. Sherman

Indoor Environment Program
Applied Science Division
Lawrence Berkeley Laboratory
University of California
Berkeley, California, 94720

November 1988

ABSTRACT

Tracer gas techniques are becoming widely used to measure the ventilation rates in buildings. As more detailed information is required for both energy and indoor air quality purposes, researchers are turning to complex, multizone tracer strategies. Both single gas and multiple gas techniques are being utilized, but only multigas are capable of uniquely determining the entire matrix of air flows. Because of the inherent limitations in the ability to estimate zonal concentrations, estimates of multizone air flows are highly imprecise for real buildings. However, exogenous information concerning physical constraints can allow a greatly improved estimate and interpretation of results if combined with measured data. This report describes techniques for improving tracer-gas derived ventilation data using physical knowledge about the system under study.

Keywords: Ventilation, Infiltration, Tracer Gas, Multizone Measurement Techniques, Error Analysis, Uncertainty

---

This work was supported by the Assistant Secretary for Conservation and Renewable Energy, Office of Building and Community Systems, Building Systems Division of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098

INTRODUCTION

Tracer gasses are used for a wide range of diagnostic techniques including leak detection[1,2] and atmospheric tracing.[3] One application which has had a resurgence in the last decade is the use of tracer gasses to measure ventilation (i.e., air flow) in buildings.[4] Ventilation is an important process in buildings because of its impact on both energy requirements and indoor air quality— both of which are topics of concern to society. Measurement of the tracer gas concentration and source emission combined with conservation laws allows a quantitative determination of the tracer transport mechanism (i.e., a measurement of the air flow).

The vast majority of the ventilation measurements made to date have involved a single-tracer gas deployed in a single zone. This technique has proven very useful for buildings which may be treated as a single zone (e.g., houses) and for more complex buildings in which there are isolatable sub-sections. However, as the need to understand more complex buildings has grown, tracer gas techniques that are able to treat multiple zones have been developed.[5] Multizone techniques recognize that not only does air flow between the outside and the test space, but that there are air flows between different parts (i.e., zones) of the test space and, in the complete case, they are able to measure these flows.

Because of the multiple sources of randomness associated with multizone tracer gas studies, precision may be relatively poor. However, a user of such tracer gas techniques has more information available to him than is contained within the data alone. This *a priori* information can greatly improve the accuracy and precision of the measurement if properly combined with the data. This report will endeavor to show how to combine prior information with the data to get an *a posteriori* set of air flow estimates and associated errors that improves upon the data.

BACKGROUND

The continuity equation expresses the conservation of tracer gas. In a general multizone environment, a *matrix* form of the continuity equation must be used:

$$\mathbf{V}\cdot\dot{\mathbf{C}}(t) + \mathbf{Q}(t)\cdot\mathbf{C}(t) = \mathbf{S}(t) \tag{1.1}$$

For every zone of the system there will be a row in both the concentration and source-strength matrices. For every unique tracer gas there will be a column in those matrices. If there are N zones, the volume and air flow matrices will be square matrices of order N and the continuity equation can be rewritten with explicit indices:

$$\sum_{j=1}^{N} \left[V_{ij}\dot{C}_{jk}(t) + Q_{ij}(t)C_{jk}(t) \right] = S_{ik}(t) \tag{1.2}$$

If there are as many tracer species as there are zones, the problem is called *complete* and there will be an exact answer; we shall focus our attention to the complete problem and therefore assume that all of the matrices are square.

As Roulet[6] points out, the continuity equation is a *mass* balance equation and serious errors can result if it is used as a *volume* balance equation unless proper precautions are taken. Accordingly, the concentrations are expressed in mass of tracer per unit volume to assure correctness even when the density of air varies from zone to zone (e.g., if the zones are at different temperatures).

Interpretation of Matrix Elements

Eq. 1 contains measured data and derived quantities. The measured data are the flows and concentrations of each tracer gas in each zone. Specifically, C_{ij}, $\dot{C}_{ij}$, and S_{ij} all represent the respective value of the jth tracer gas in the ith zone.

The volume matrix can either assumed to be independently determined or derived from the measured data. It is usually assumed (and will be herein) that the volume has been exogenously determined. For most practical purposes the volume matrix can be assumed diagonal with the individual zone volumes as the entries. If, however, there is *short circuiting* of the tracer source from one zone to another, it can manifest itself as an off-diagonal volume element, but the sum of each column must be equal to the (effective) physical volume of the zone.

The interpretation of the air flow matrix requires a bit more explanation. The diagonal elements, Q_{ii}, represent the total flow out of that zone to all other zones and should have positive sign. The off-diagonal elements represent the flows between zones; specifically, $-Q_{ij}$ is flow from the jth zone to the ith zone. Since the flow from the jth zone to the ith zone can be different from the flow from the ith zone to the jth zone, this matrix will in general not be symmetric.

The flow matrix explicitly contains information about flows between measured zones and the total flow. If there are flows to zones other than those being measured (e.g., outside), the sum of some rows and columns of the flow matrix will be positive; and system is said to be *open*. If all zones of the building are monitored these flows to ‘‘elsewhere’’ are attributed to air exchange with the outside.

ESTIMATION OF FLOWS FROM MEASURED DATA

Inversion of eq. 1 is a straightforward mathematical problem:

$$\mathbf{Q}(t) = \left[\mathbf{S}(t) - \mathbf{V}\cdot\dot{\mathbf{C}}(t) \right]\cdot\mathbf{C}(t)^{-1} \tag{2}$$

If there were no uncertainty in the measured data (i.e., the concentrations and source strength), this inversion would give the correct (and only) answer. In any real experiment, however, there will be uncertainty in the measurements due either to instrumentation errors or other random processes. Such uncertainty can be described by a probability distribution as to where the true value lies. Tarantola[7] gives an excellent discourse on the issues related to the general problem of extracting model parameters from measured data.

The covariance of the data can be calculated if the uncertainties in the measured concentrations and source strengths are known;[8] all of the data covariances* used herein are so calculated, but the results of this report can be used howsoever the covariance is determined. The remainder of this report assumes that the errors can be assumed to be Gaussian. This common assumption may not be strictly true for a variety of reasons (e.g., the flow and concentration values are positive definite, mixing is not a Gaussian process, etc.). For most common applications, however, the assumption is unlikely to lead to significant errors and we will use it. If

* The reader should be careful to note the number of dimensions used in the matrix notation. The air flows are naturally treated as a matrix (i.e., tensor of rank 2) which implies that the covariance "matrix" is really of rank 4. To compare with more standard treatments, the air flows *could* be considered as a vector and the covariance as a normal matrix. As long as it is realized that the covariance matrix has twice the dimensions of the air flows, the matrix notation will be left general.

the covariance of the flows are known and can be assumed Gaussian, the probability distribution for the true value is as follows:

$$\Psi_o(\mathbf{Q}|\,\mathbf{Q_d}) = \xi_o(\mathbf{Q_d}) \;\; e^{-\frac{1}{2}\|\,\mathbf{Q},\mathbf{Q_d}\|^{2}} \tag{3}$$

where subscript "d" implies that the quantity is calculated directly from the data. The normalization for the probability is as follows:

$$\xi_o(\mathbf{Q_d}) = \left[(2\pi)^{N^2}\left|\,\boldsymbol{\sigma}(\mathbf{Q_d})\right|\;\right]^{-\frac{1}{2}} \tag{4.1}$$

and the determinant:

$$|\,\boldsymbol{\sigma}| \equiv \textit{Determinant of the Covariance Matrix} \tag{4.2}$$

The norm used in the above equations,

$$\|\,\mathbf{Q},\mathbf{Q_d}\|^{2} \equiv \left[\mathbf{Q}-\mathbf{Q_d}\right]^{t}\cdot\boldsymbol{\sigma}(\mathbf{Q_d})^{-1}\cdot\left[\mathbf{Q}-\mathbf{Q_d}\right] \tag{5}$$

represents a normalized distance between two points using the covariance as the weighting (i.e., the metric of the space). (This square may be familiar to the reader as a χ^2 variable.)

It should be noted that the mean, median, and maximum likelihood estimator (i.e., the mode) of the distribution are all equal to the point value as calculated by the inversion of eq. 1 (i.e., $\mathbf{Q_d}$).

Incorporation of Prior Estimates

It is quite often the case that we have some knowledge about the result that does not come from the current data. Such a priori information is called prior knowledge or more simply referred to as the "priors". Two common examples of such prior knowledge would be an independent measurement of (some of) the same quantities, or some physical knowledge about a particular flow. If this knowledge can be expressed as mean set of air flows, $\mathbf{Q_p}$, with (Gaussian) covariances, $\boldsymbol{\sigma}_\mathbf{p}$, we can combine our measured value with our prior knowledge to improve our estimate of the true value:

$$\mathbf{Q} = \boldsymbol{\sigma}\cdot\left[\boldsymbol{\sigma}_\mathbf{d}^{-1}\cdot\mathbf{Q_d} + \boldsymbol{\sigma}_\mathbf{p}^{-1}\cdot\mathbf{Q_p}\right] \tag{6.1}$$

$$\boldsymbol{\sigma} = \left[\boldsymbol{\sigma}_\mathbf{d}^{-1} + \boldsymbol{\sigma}_\mathbf{p}^{-1}\right]^{-1} \tag{6.2}$$

Even if the prior knowledge is very uncertain, its effect can only be to improve our estimate of the true values, provided we know how uncertain the prior knowledge is. This improvement can be especially useful when the problem is poorly conditioned and one or more flow elements may be extremely uncertain. Any relationship that can be expressed linearly can be reflected in the prior. If little prior knowledge is available for a particular element, any reasonable value may be used— provided that its variance is large enough to cover the bulk of its allowed range.

The prior covariance and the final covariance can be combined to give a resolution operator[7] which describes the quantities that are well-resolved by the data and those that are not. The resolution matrix is defined as follows:

$$\mathbf{R} \equiv \mathbf{I} - \boldsymbol{\sigma}_{\mathbf{p}}^{-1} \cdot \boldsymbol{\sigma} \tag{7}$$

$$\mathbf{R} \cdot \mathbf{Q}_\lambda = r_\lambda^2 \, \mathbf{Q}_\lambda \tag{8}$$

The eigenvectors of this matrix are the linear combination of the parameters which are independently resolved by the data. The eigenvalues represent how well the data resolves those vectors: an eigenvalue of zero means that all the information about that combination came from the prior knowledge (i.e., no information in the data); while an eigenvalue of unity means that the data completely determined that combination of parameters. r_λ, the square root of the eigenvalue, plays the role of a correlation coefficient in a multilinear regression in that it determines how well a quantity ($\mathbf{Q}_\lambda$) is determined by the data (as represented by $\boldsymbol{\sigma}_{\mathbf{d}}$).

Sometimes it is stated that no prior information was known about the flows. But if we interpret "no prior information" to mean that there is a large (uncorrelated) variance on each term and that that variance is constant for all parameters we can reduce the problem to finding the eigenvalues and eigenvectors of the initial covariance:

$$\boldsymbol{\sigma}_{\mathbf{d}} \cdot \mathbf{Q}_\lambda = \sigma_\lambda^2 \mathbf{Q}_\lambda \tag{9}$$

These eigenvalues represent the variances of the (uncorrelated) eigenvectors. This type of principal value decomposition is used in analysis of errors, but often ignores the assumptions stated above.

There are other approaches to the analysis of error for multizone tracer problems. Roulet[6], D'Ottavio,[9] and Walker[10] have all proposed methods based on the condition number of air flow and concentration matrices. These methods do not in fact estimate the uncertainty of the air flow matrix, but rather they set bounds on specific errors. As such, they may be useful in estimating errors for the incomplete problem, but are not as powerful as the principal value methods described herein.

PHYSICAL KNOWLEDGE OF THE SYSTEM

The previous sections do not contain all of the information that is known about the system of equations. The flow matrix does not even contain any elements relating to flows to or from outside— which are usually the flows of most interest. To properly interpret the results more physical knowledge is needed.

Eq. 1.2 is an *open* set of equations; that is, there are flows that can go to and from "elsewhere", where "elsewhere" is usually interpreted to mean outside. These flows are inferred by assuming that there are no unaccounted for sources or sinks of tracer and that the volume of air flowing in and out of a zone is equal.

Augmentation of the Matrices

We can make these physical assumptions more explicit by augmenting the matrices by an additional (zeroth) tracer to account for the conservation of air and zone to account for outside flows. Thus Q_{i0} represents the flow from outside to the *i*th zone, and Q_{0j} represents the flow to outside from the *j*th zone, and Q_{00} represents the total flow to all zones from outside.

The outside zone is different from the other zones in several ways. The outside zone must supply closure to the system so that there is no net flow of air or tracer into or out of the system. For this to be true the following conditions must be met:

$$\sum_{i=0}^{N} S_{ik} = 0 \qquad k=0....N \tag{10.1}$$

which in turn implies that

$$\sum_{i=0}^{N} V_{ij} = 0 \qquad j=0....N \tag{10.2}$$

and, therefore,

$$\sum_{i=0}^{N} Q_{ij} = 0 \qquad j=0....N \tag{11}$$

These three expressions then serve as defining relations for the zeroth row of there respective matrices. The zeroth row of the concentration matrix may contain any background concentration of tracer gas.

$$C_{0k} \equiv outside\ concentration\ of\ gas\ k \qquad k=0 \cdots N \tag{12}$$

The zeroth column of the concentration matrix is the density of air in the zones:

$$C_{j0} \equiv \rho_j \qquad j=0 \cdots N \tag{13.1}$$

Since the there is no addition of air to any zones,

$$S_{i0} \equiv 0 \qquad i=0 \cdots N \tag{13.2}$$

the air flow matrix must meet the following criterion:

$$\sum_{j=0}^{N} Q_{ij}\rho_j = -\sum_{j=0}^{N} V_{ij}\,\dot{\rho}_j \qquad i=0 \cdots N \tag{14}$$

Note that if the density of air is invariant and equal in every zone, this relationship is the transpose analogue of eq. 11.

Finally, in order to continue to meet eq. 1 in the augmented style, the remaining volume terms must be zero:

$$V_{i0} \equiv 0 \qquad i=0 \cdots N \tag{15}$$

Thus the outside zone is treated as a fully coupled zone with zero effective volume, but with tracer sinks.

We have then two equivalent descriptions: the set of equations defined by the eqs. 1 10 12 and 13. or the unaugmented set of equations, plus the definitions derived for the outside flows eqs. 11 14. Assuming that the time change of density in any zone is small, these flows can be rewritten as follows:

$$Q_{0j} = -\sum_{i=1}^{N} Q_{ij} \qquad j=0 \cdots N \tag{16.1}$$

$$Q_{i0} = -\sum_{j=1}^{N} Q_{ij}\,\frac{\rho_j}{\rho_0} \qquad i=0 \cdots N \tag{16.2}$$

Either method may be used to determine the flow matrix, but since the augmented matrices are larger by one dimension, their inversion may take significantly longer. Hence, for computational efficiency it is better to use the unaugmented version.

Since these equations completely determine the flows to and from outside, the probability description of the air flow matrix need not *explicitly* contain them. Accordingly, the matrix represented by the symbol **Q** should be taken to mean the *un*augmented flow matrix, with the outside flows implied by eq. 16.

Physicality Constraints

In our definition of flow matrices only certain values of individual elements represent physically meaningful values, which are reflected in the signs of the matrix elements. The physicality constraints can be summarized as follows:

$$Q_{ij} \le 0 \qquad \left\{ \begin{array}{l} i=0 \cdots N \\ i \ne j \\ j=0 \cdots N \end{array} \right. \tag{17}$$

which when included with eqs. 16 yield the following weak condition:

$$Q_{ii} > 0 \qquad\qquad i=0 \cdots N \tag{18}$$

Taken together eqs. 16 and 17 represent the physicality constraints on the air flow matrix.

These expressions for incorporating prior knowledge in the previous section are strictly true only for Gaussian distributions. In the case of these physicality constraints, we have some critical prior knowledge that cannot be expressed with a Gaussian covariance; specifically, we know that the true value must meet all of the physicality constraints. Although we cannot use the Gaussian expressions, we can modify the probability distribution to account for our knowledge:

$$\Psi(\mathbf{Q}|\,\mathbf{Q_d}) = RHO(\mathbf{Q})\ \xi(\mathbf{Q_d})\ e^{-\frac{1}{2}\left(||\ \mathbf{Q},\mathbf{Q_d}||\ \right)^2} \tag{19}$$

where RHO represents the physicality constraint,

$$RHO(\mathbf{Q}) = \begin{cases} 1 & \text{if } \mathbf{Q} \text{ is physically possible} \\ 0 & \text{otherwise} \end{cases} \tag{20}$$

and again ξ normalizes the distribution:

$$\xi(\mathbf{Q_d}) = \frac{1}{\int RHO(\mathbf{Q})\ e^{-\frac{1}{2}\left(||\ \mathbf{Q},\mathbf{Q_d}||\ \right)^2} d^{N^2}\mathbf{Q}} \tag{21}$$

Thus the (unnormalized) distribution RHO represents our prior knowledge.

Because the addition of the physicality constraints has truncated the erstwhile normal distribution, the mean, median, and mode will all have changed values. These values will no longer be equal to each other and none of them will be a totally unbiased estimator of the the true value. If we wish to characterize this distribution by a single parameter, any of the three, a priori, could be considered, but each has different consequences. Once the estimator has been found, however, the covariance can be calculated from it and the distribution:

$$\boldsymbol{\sigma} = \boldsymbol{\sigma}_{\mathbf{d}} \frac{1}{N^2} \int ||\ \mathbf{Q},\hat{\mathbf{Q}}\,||\ ^2 \Psi(\mathbf{Q}|\,\mathbf{Q_d})\, d^{N^2}\mathbf{Q} \tag{22}$$

Thus, the posterior covariance is just a simple multiple of the covariance of the original data.

The posterior variance will be minimized if the estimator is the mean of the distribution. When the initial point, $\mathbf{Q_d}$, is allowed and far from the (physicality) boundary, this integral will be unity and the two covariances are the same; as the initial point approaches the boundary the posterior covariance gets smaller; as the initial point moves into the disallowed area, the posterior covariance gets quadratically larger. The expectation value of such posterior covariance (integrated over the distribution of initial points given a true value) is the same as the prior covariance, which suggests that the integral above plays the role of a chi-squared statistic. Unfortunately, it does not follow a chi-squared distribution. Shapiro[11] has shown that our type of distribution can be described by one which is a combination of chi-squared distributions having degrees of freedom up to the total number (i.e., N^2 in our case).

The fact that the mean has minimum variance might suggest that it is the estimator of choice. For our purposes, however, the mean has several disadvantages (which will be left unproven) that make it unsuitable for use as the point estimate. The mean of our (truncated) distribution is a biased measure (i.e., on average it will tend to predict a point estimate that is slightly further away from the physicality limits than the true value). The bias is biggest when the true value is nearest these limits. Since we expect that (for the interzonal flows) the true values will often be at the limits (i.e., there will be no interzonal flow between some zones), the mean is an inappropriate point estimate.

The median is sometimes considered an estimator because it is more robust. That is, it is less sensitive to low probability events. The median, however, is always a more biased estimate of the true value for our case than is the mean. Since the robustness of the median is not of significant usefulness to justify the increased bias, the median is not an appropriate point estimate.

In contradistinction to the mean and median, the maximum likelihood indicator has several advantages as an estimator. By definition it is the *most likely* point to be the true value. Although biased, the maximum likelihood indicator is a less biased indicator than the mean for our case. Furthermore, when the initial point (i.e., $\mathbf{Q_d}$) is physically allowed, that initial point *is* the maximum likelihood indicator. Since for most tracer applications it is the point estimate rather than its variance which is most important, we choose the maximum likelihood estimator as our point estimate of the true value of the air flows.

When the initial point is allowed no further calculation to get the maximum likelihood indicator in necessary. When the original point is physically prohibited, the determination of the estimator can be reduced to a minimization problem with bounded values where we minimize the norm:

$$\hat{\mathbf{Q}}: \quad || \, \hat{\mathbf{Q}}, \mathbf{Q_d} \, ||^2 \quad \textit{is a minimim} \tag{23}$$

If Gaussian priors are available as well as the physicality constraints, they should be applied before the physicality procedure. If the new point is still disallowed then the minimization should be over both initial point and prior estimate:

$$\hat{\mathbf{Q}}: \quad || \, \hat{\mathbf{Q}}, \mathbf{Q_p} \, ||^2 + || \, \hat{\mathbf{Q}}, \mathbf{Q_d} \, ||^2 \quad \textit{is a minimim} \tag{24}$$

Example

As an example of this technique we consider the dataset presented by D'Ottavio[9] in which the matrices were augmented. The augmentation procedure was not complete, but was equivalent to the one presented herein, assuming all of the air densities were constant and equal (and arbitrarily set to unity) and the concentrations were invariant. Then a matrix error propagation method that assumed small, normally distributed errors was used to find the uncertainties in the flows. Although the technique is different from the general technique of ref. 8, the results are equivalent for the special case of the data and are displayed in table 1.

TABLE 1: Example Air Flows and Uncertainties for PFT Dataset [m^3/hr]

$Q_{ij} \pm \sigma_{Q_{ij}}$	1	2	3	Outside
1	667±107	-314±64	15±25	368±61
2	-132±43	454±52	-212±33	110±33
3	-17±5	-23±6	293±43	254±37
Outside	518±92	118±69	97±42	733±59

The errors can be calculated from the original reference. The calculation of the covariance was not done by the authors, but was done in a separate report[8]. Using our initial (i.e., uncorrected) values of the air flows and covariance matrix we can find the linear combinations of flows that make up the principal components of this data and display them in table 2.

Table 2: Coefficients of Principal Components of Measured Data

#	σ_λ[m^3/hr]	1,1	2,1	3,1	1,2	2,2	3,2	1,3	2,3	3,3
3	10	*-0.76*	*-0.65*	-0.03	-0.03	0.04	-0.00	-0.06	0.01	-0.01
6	37	*-0.65*	*0.76*	-0.03	-0.01	-0.04	0.04	-0.06	-0.03	0.00
7	46	-0.00	0.01	*0.59*	*-0.55*	0.21	0.54	0.07	0.07	0.06
8	61	-0.01	0.01	*0.54*	*0.53*	0.38	-0.09	-0.31	-0.42	-0.03
2	5	-0.03	-0.06	0.19	0.24	**-0.73**	0.34	0.28	-0.38	-0.20
4	16	-0.01	0.02	0.38	-0.39	-0.23	**-0.65**	-0.07	0.01	-0.47
9	124	-0.10	0.03	0.18	0.23	0.27	-0.22	**0.87**	0.17	0.00
5	26	-0.02	-0.01	0.34	0.37	-0.30	0.03	-0.22	**0.77**	0.14
1	2	-0.02	-0.01	0.18	-0.16	-0.25	-0.33	0.05	-0.23	**0.85**

The rows are ordered in an approximately diagonal fashion for clarity. The eigenvalue numbers indicate order of increasing variance, so that the lower numbers are the most well determined combinations and the higher numbers are the least well determined. The coefficients have been normalized to unity. As indicated by the **bold** entries certain vectors are dominated by a single air flow value; as indicated by the *italics* certain pairs of vectors are dominated by pairs of air flows.

There are several interesting observations one can make. The most well determined combination (i.e., eigenvalue #1) is dominated by the 3,3 air flow (i.e., the total flow in or out of zone 3). Thus is appears that this combination is much more well determined (cf. 2.2 m^3/hr) than the precision on that element (from table 1) would suggest (cf. 43 m^3/hr). In a similar way the last entry (i.e., #9) which represents the least well determined combination of air flows is dominated by the 1,3 value and is more poorly determined (cf. 124 m^3/hr) than its variance would indicate (43 m^3/hr). Such a result is not surprising— noting that that element was the one which came out with a physically impossible result.

There are two pairs of rows (and columns) which are dominated by a nearly equal pair of values— indicating that the sum and difference of these two air flows is a principally determined quantity. With the exception of the #3, all of the eigenvectors that involve zone 1 have large variances; such a result may indicate a problem with the measurements in that zone.

We can take this example further by putting the physicality constraints on the point estimates. If we integrate the probability distribution over the allowed space, we find that only 31% of the initial distribution is in the physically permitted space.

In this example Q_{13} is physically disallowed. We can use the minimization technique to find the best possible solution. Using the covariance and the physicality constraints the adjusted results become the following:

TABLE 3: Fixed Air Flows and Uncertainties for PFT Dataset [m^3/hr]

$Q_{ij} \pm \sigma_{Q_{ij}}$	1	2	3	Outside
1	653±94	-291±56	0±22	362±54
2	-130±38	448±46	-206±29	113±29
3	-17±4	-23±5	292±38	253±33
Outside	506±81	134±61	87±37	727±52

The new point is, of course, physically allowed so that the offending element has been moved to the boundary. To do this with minimal change in the norm required that some of the other elements be adjusted also. The covariance calculated at this new point is different for two reasons: 1) the central value is slightly different and 2) the integral in eq. 22 induces a scale factor based on the minimization. The first reason is a small shift, but the scale factor for this dataset is approximately 0.9.

TIME-SERIES DATA

The preceding sections have described methods to estimate air flows from a single set of measured data. Many of the tracer gas systems currently in use measure the concentrations and flows at a high data rate; that is, there are many measurements in the time it takes for the system (i.e., the air flows) to change significantly. In such a case, the data contains redundant information, which can be used to improve the estimate of the underlying air flow.

Physically we know that the flow values are correlated in time, and we can assume that we can estimate a (Gaussian) correlation time for the system— τ_Q. The most straightforward approach to solving the time series problem would be to do a fit for each time point (as described above) combined with a simultaneous correlation in time for all the points. However, since in such a global approach the number of dimensions in the fit grows very large, the computing requirements become unreasonable. (Requirements typically go as the cube of the number of dimensions.) Furthermore, such a global procedure is *acausal*; that is, values at any time are related to events that happened both before and *after* the event. We, therefore, would prefer an analysis method that is causal, local and contains fewer dimensions.

If we assume that over some time period, τ_Q, the underlying system does not change much, we can then integrate the continuity equation (eq. 1) over this period and treat the air flows as constant. The integrated continuity equation will have a significantly smaller (data) covariance than the instantaneous one. The longer the integration time is the more precise the determination of the air flows will be. One must be careful, however, not to make the time too long or the assumption of constant air flows will break down and a bias will be introduced. The *passive ventilation* measurement technique suffers from this bias.[12] The trade-off between precision and bias in the selection of the integration time constant requires some prior knowledge about the system; however, for houses without mechanical systems, this time is typically on the order of one hour.

Because we know that the underlying air flows are smoothly varying causal functions of time, we can use a prior (i.e., previous in time) estimate of the air flows to improve a current estimate. To do so we must estimate an upper bound to the covariance between two air flow values separated by an integration time, and then use the prior-knowledge technique to find a new estimate. A very reasonable assumption is that the air flows change by much less than one air change rate in a correlation time. Thus if two measurements are made at a time δt apart, the prior covariance should be as follows:

$$\boldsymbol{\sigma}_{\mathbf{p}} \ll \left[\frac{\delta t^2}{\tau_Q^4} \right] \mathbf{V}\mathbf{V} \tag{25}$$

Usually we assume the volume matrix to be diagonal and therefore the covariance matrix will be diagonal and has a very strong dependence on the correlation length, τ_Q.

If more specific knowledge of the time behavior of air flows is known, a more detailed prior covariance than the above equation could be used. However, in most cases there is little to be gained by such a procedure.

The time series analysis technique can be summarized as follows:

- *INTEGRATE* the continuity equation over the correlation time (τ_Q) at a number of overlapping points separated by a convenient time ($\delta t < \tau_Q$) spacing.
- *INVERT* the integrated data to get an estimate of the flow values ($\mathbf{Q_d}$) calculate the covariance of the data ($\boldsymbol{\sigma}_{\mathbf{d}}$), and calculate outside flows.
- *COMBINE* this initial estimate with the previously calculated time point ($\mathbf{Q_p}$) using a prior covariance ($\boldsymbol{\sigma}_{\mathbf{p}}$) based on the correlation length.
- *ADJUST*, if necessary, these values ($\mathbf{Q}$) to meet the physicality constraints by minimizing both the norm calculated from the initial point and the prior value using the appropriate covariances.
- *CALCULATE* the final covariance ($\boldsymbol{\sigma}(\hat{\mathbf{Q}})$)based on the final value.
- *REPEAT* for every time point in the dataset
- *POST PROCESS* the data for presentation or reduction. Weighted averaging may be accomplished using the covariance. The data may be filtered or smoothed to further reduce noise or unwanted frequencies without adding bias.

Time-Series Example

LBL has recently completed a MultiTracer Measurement System (MTMS),[13] that uses multiple tracer gasses in a fully automated manner to measure flows and concentrations for the purpose of determining air flows in a multizone environment. The analysis technique described herein was used to estimate the air flows from the data measured with MTMS.

As an example we have chosen a two-story house situated in the Seattle area of the state of Washington. Figure 1 is a plot of air flow data calculated by integrating the data over a one hour period but no adjustment or incorporation of data was done on it. There are an entire set of air flows which may be graphed, but we have elected to show only the ones which are flows to the (upstairs) bedroom zone. The uncertainties are typically in the range of 1-10 m^3/hr. As can be seen the data is very unstable; such behavior is not surprising as there is unsteady mixing of the tracer gasses in this house. Many of the points are clearly unphysical.

The house is insulated and has forced air (electric heat-pump) heating. This two story house has one third of the lower story taken up by an (unconditioned) garage; the main living area is upstairs. The floor plan is open with a wide, open stairway between the two floors. Each room has a heating register and the return is located in *living area*.

Figure 2 is the same dataset after all of the steps in the analysis are completed. Most of the points needed to be adjusted and in some cases the adjustment was highly significant; however, when the posterior covariance was used to calculate a weighted average, the uncertainties are not changed significantly from the unimproved data. In general the data is both more accurate and more precise.

From the finer results in figure 2 it is possible to ascertain some of the behavioral effects going on in this house. For example, in the morning the heating system comes on after set back and presumably interior doors are opened and windows shut, while at bedtime the opposite happens. This shift can most readily be seen in the January 15 data in which the nightime flow between the two upper floor zones is increased at bedtime, while the flow from the lower zone is decreased as the forced air system no longer distributes the air between zones.

CONCLUSION

Point estimates of air flows are often desired for determining energy and pollutant flows in multizone buildings. Because of the high degree of correlation between different components, the analysis of multizone air flows from tracer concentration and flow data is a difficult task. The simple analysis techniques, which are typically used to analyze such data, do not take into account the high degree of correlation and therefore may not provide a good picture of the situation.

This report has demonstrated that the estimates and their interpretation can be significantly improved by using the information contained in the covariance matrix. The following recommendations use the covariance to improve the point estimate and its interpretation:

- A principal component decomposition of the covariance matrix gives a better indication of the precision of the point estimate than do the individual variances.
- If prior information is available it should be incorporated into the estimate. The resolution matrix can then be used to determine how much the data determines the air flow parameters.

- If the point estimate is physically impossible, the norm of the difference between that point and the physical possible estimate should be minimized subject to the physicality constraint. The final covariance matrix can then be scaled based on the fit.

For time-series data:

- A correlation time should be chosen based on knowledge of the system being measured.
- For each desired output point the continuity equation should be integrated over the correlation time to maximize precision and minimize bias.
- The previous time series point should be used as a prior in the calculation of the current value.

Once the best estimate of the air flow matrix has been calculated, its covariance can be used to estimate its uncertainty. The eigenvalue techniques presented herein are superior to condition number error estimates because they can be used to 1) determine which linear combination of air flows can be determined independently, 2) how well determined are those combinations and 3) how much of the determination is due to the measured data and how much is due to prior knowledge about the system.

REFERENCES

1. P.H. Singer, "Helium Leak Detection in Vacuum Systems and IC Packages", *Semiconductors International* **10 (11),** pp. 51-5 (1987)

2. R., McCullough, "Leak Testing", *ASTM Standardization News,* **10 (11)**, pp. 32-3 (1982)

3. Ferber, G.J., Telegadas, K., Heffter, J.L., et al., "Demonstrations of a Long-Range Atmospheric Tracer System Using Perflourocarbons," BNL-33846, Brookhaven National Laboratory, (1983)

4. Lagus P., Persily, A.K., "A Review of Tracer Gas Techniques for Measuring Airflows in Buildings," *ASHRAE Transaction* **91**(IIB) p1075-1087, (1985)

5. Harrje, D.T., Dutt, G.S., Bohac, D.L., "Documenting Air Movements and Infiltration in Multicell Buildings Using Various Tracer-gas Techniques," *ASHRAE Transaction* , **91(II) (1985).**

6. Roulet, C.A., Compagnon, R, "Multizone Tracer Gas Infiltration Measurements Interpretation Algorithms for Non Isolthermal Cases, *Building & Environment,***24** (3) pp. 221-227 1989.

7. Tarantola, Albert, *Inverse Problem Theory,* Elsevier, New York, (1987)

8. Sherman, M.H., "Uncertainty in Air Flow Calculations Using Tracer Gas Measurements," To be published in *Building and Environment.*

9. D'Ottavio, T.D. et. al., "Error Analysis Techniques for PerFluorocarbon Tracer Derived Multizone Ventilation Rates," *Building & Environment*, **23**(4), pp. 187-194, 1988.

10. Walker, R.R., *Interpretation and Error Analysis of Multi-tracer Gas Measurement to Determine Air Movements in a House,* Proceedings of the 6th AIVC Conference "Ventilation Strategies and Measurement Techniques", Air Infiltration and Ventilation Centre, U.K. 1985

11. A. Shapiro, "Asymptotic Distribution of Test Statistics in the Analysis of Moment Structures under Inequality Constraints," *Biometrika*, **72** (I) pp. 133, (1983)

12. Sherman, M.H., "Analysis of Errors Associated with the Passive Ventilation Measurement Technique," *Bldg. & Environ.* (1989). Lawrence Berkeley Laboratory Report LBL-23088 (1988)

13. D. Dickerhoff, M. Sherman, "Technical Description of the Multigas Tracer Measurement System," In preparation.

NOMENCLATURE

C	Instantaneous tracer gas concentration [kg/m^3]
$\mathbf{C}$	Multizone tracer gas concentration matrix [kg/m^3]
$\mathbf{I}$	Identity Matrix [-]
N	Number of zones [-]
Q	Ventilation [m^3/h]
$\mathbf{Q}$	Ventilation matrix [m^3/h]
$\mathbf{Q_d}$	Ventilation matrix from measured data [m^3/h]
$\mathbf{Q_p}$	Ventilation matrix from prior information [m^3/h]
$\hat{\mathbf{Q}}$	Point estimate of ventilation matrix [m^3/h]
$\mathbf{R}$	Resolution matrix [-]
r	Correlation coefficient [-]
S	Instantaneous source strength of tracer gas [m^3/h]
$\mathbf{S}$	Multizone tracer source strength matrix [m^3/h]
t	Time [h]
δt	Time difference between measurements [h]
τ_Q	Correlation (Integration) time [h]
V	Volume [m^3]
$\mathbf{V}$	Zone volume matrix [m^3]
ρ	Density of air in a zone [kg/m^3]
RHO	(Unnormalized) Distribution of physically allowed values [-]
σ	Standard deviation of an air flow [m^3/hr]
$\boldsymbol{\sigma}$	Covariance matrix of air flows [m^3/hr]^2
$\boldsymbol{\sigma}_\mathbf{d}$	Covariance matrix from measured data [m^3/hr]^2
$\boldsymbol{\sigma}_\mathbf{p}$	Covariance matrix from prior information [m^3/hr]^2
ξ	Probability normalization [-]
i,j,k	Indices indicating zone [$0,1 \cdots N$]
λ	Index indicating eigenvalue [$1 \cdots N^2$]

LIST OF FIGURES